In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```
df = pd.read_csv("AirPassengers.csv")
```

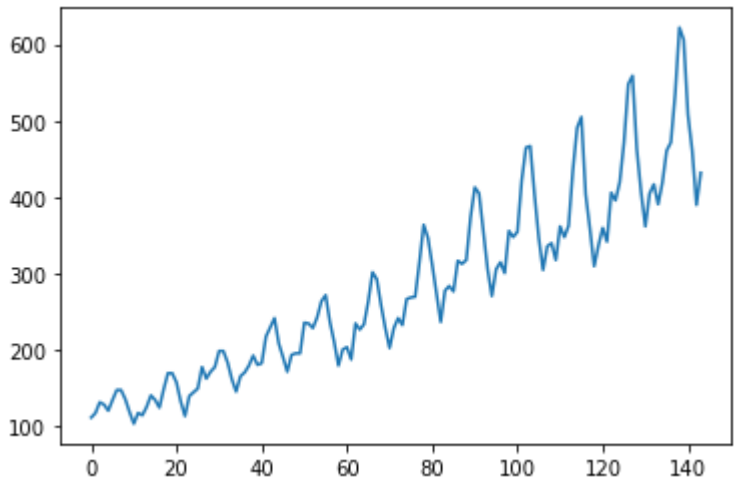
In [3]:
```
df.head(10)
```

Out[3]:

	Month	#Passengers
0	1949-01	112
1	1949-02	118
2	1949-03	132
3	1949-04	129
4	1949-05	121
5	1949-06	135
6	1949-07	148
7	1949-08	148
8	1949-09	136
9	1949-10	119

In [4]:
```
df["#Passengers"].plot()
```

Out[4]: <AxesSubplot:>



In [5]:
```
df["diff_shift_1"] = df["#Passengers"] - df["#Passengers"].shift(1)
```

In [6]:
```
df["diff_shift_1"]
```

Out[6]:
```
0       NaN
1       6.0
2      14.0
3      -3.0
4      -8.0
       ...
139   -16.0
140   -98.0
141   -47.0
142   -71.0
143    42.0
Name: diff_shift_1, Length: 144, dtype: float64
```

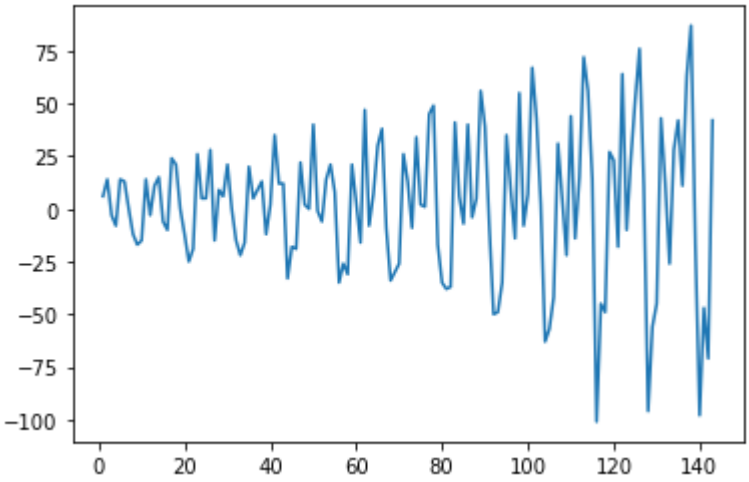
In [7]:
```
df.head(3)
```

Out[7]:

	Month	#Passengers	diff_shift_1
0	1949-01	112	NaN
1	1949-02	118	6.0
2	1949-03	132	14.0

In [8]:
```
df["diff_shift_1"].plot()
```

Out[8]:
```
<AxesSubplot:>
```



In [9]:
```
from statsmodels.tsa.stattools import  adfuller
```

In [10]:
```
adfuller((df["diff_shift_1"]).dropna())
```

Out[10]:
```
(-2.8292668241699888,
 0.054213290283826945,
 12,
 130,
 {'1%': -3.4816817173418295,
  '5%': -2.8840418343195267,
  '10%': -2.578770059171598},
 988.5069317854084)
```

In [11]:
```
def adf_test(series):
    result = adfuller(series)
    print("p - values : {}".format(result[1]))
    if result[1]<=0.05:
        print("strong evidence against the null hypothesis,reject null hypothes
    else:
        print("weak evidence against null hypothesis, indicating that the data
```

In [12]:
```
adf_test(df["diff_shift_1"].dropna())
```

```
p - values : 0.054213290283826945
weak evidence against null hypothesis, indicating that the data is non -stati
onary
```

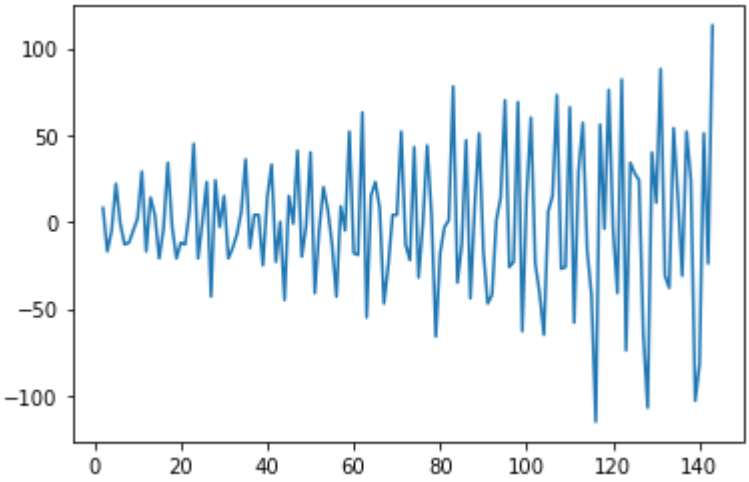
In [13]:
```
df["diff_shift_2"] = df["diff_shift_1"] - df["diff_shift_1"].shift(1)
```

In [14]:
```
df["diff_shift_2"]
```

Out[14]:
```
0        NaN
1        NaN
2        8.0
3      -17.0
4       -5.0
       ...
139   -103.0
140    -82.0
141     51.0
142    -24.0
143    113.0
Name: diff_shift_2, Length: 144, dtype: float64
```

In [15]:
```
df["diff_shift_2"].plot()
```

Out[15]: <AxesSubplot:>



In [16]:
```
adf_test(df["diff_shift_2"].dropna())
```

```
p - values : 2.732891850014085e-29
strong evidence against the null hypothesis,reject null hypothesis, indicatin
g that data is stationary
```

In [17]:
```
from statsmodels.tsa.ar_model import AutoReg
```

In [18]:
```
dff = df["diff_shift_2"].dropna()
dff.shape
```

Out[18]: (142,)

In [19]:
```
train = dff[:len(dff)-7]
```

In [20]:
```
train.shape
```

Out[20]: (135,)

In [21]:
```
test = dff[len(dff)-7:]
```

In [22]:
```
test.shape
```

Out[22]: (7,)

In [23]:
```
model = AutoReg(df["diff_shift_2"].dropna(),lags=1).fit()
```

```
C:\Users\User39\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: An unsupported index was provided and will be ignored wh
en e.g. forecasting.
  self._init_dates(dates, freq)
```

In [24]:
```
model
```

Out[24]:
```
<statsmodels.tsa.ar_model.AutoRegResultsWrapper at 0x262bd063820>
```

In [25]:
```
pred = model.predict(start =136,end=142)
```

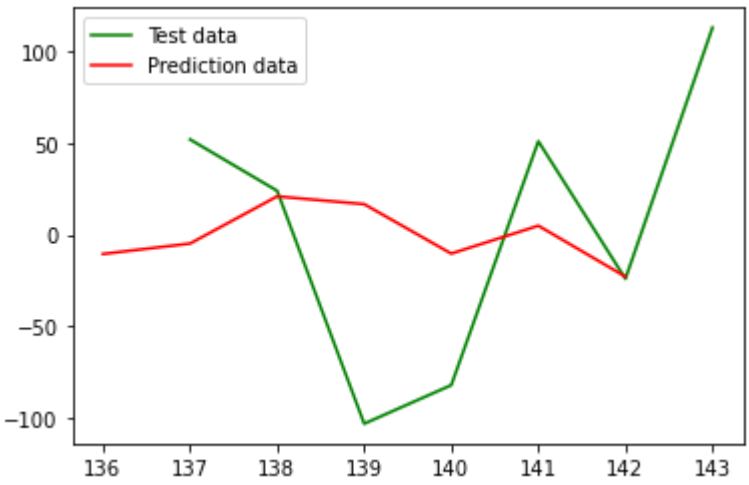
```
C:\Users\User39\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:834: ValueWarning: No supported index is available. Prediction results will
be given with an integer index beginning at `start`.
  return get_prediction_index(
```

In [26]:
```
len(dff)-1
```

Out[26]:
```
141
```

In [27]:
```
plt.plot(test,label="Test data",color='g')
plt.plot(pred,label="Prediction data",color='r')
plt.legend()
```

Out[27]:
```
<matplotlib.legend.Legend at 0x262bd0c1e20>
```



In [28]:
```
from sklearn.metrics import mean_squared_error
```

In [29]:
```
rmse = np.sqrt(mean_squared_error(test,pred))
```

In [30]:
```
rmse
```

Out[30]:
```
86.90562287963606
```

In [ ]:
```

```